3. For a given sample size $n$, consider observing $y_i$ from a sample linear model without an intercept, with normal i.i.d. errors and $x_i = i/n$. For $\hat{\beta}_{MLE}$ show that:

(a) Without calculating an analytic form of the variance, argue that $\hat{\beta}_{MLE}$ is a consistent estimator of $\beta$.

(c) Assume that $\sigma$ is equal to 2 and known. Find the smallest $n$ such that the $z$-test for the null hypothesis $H_0: \beta = 0$ will yield a $p$-value less than 0.05 when the true $\beta$ is greater than 1.

**Solution.** There are a few ways of doing this, but they all essentially boil down to showing that

$$\sum_{i=1}^{n} x_i = (1/n)^2 \sum_{i=1}^{n} i^2 \to +\infty.$$ 

That is, $\sum_{i=1}^{n} i^2$ converges to infinity faster than $n^2$. One method is to use the following inequality:

$$\sum_{i=1}^{n} i^2 \geq \int_{0}^{n} x^2 dx = n^3/3,$$

which is apparent if one uses a geometric interpretation of sums and integrals.

Another method considers only those $i$ in the sum $\sum_{i=1}^{n} x_i^2$ for which $x_i \geq 1/2$. There are at least $n/2$ of these and so the sum can be lower bounded by $n^3/8$.

(c) Under the null hypothesis of $\beta = 0$, $\frac{\hat{\beta}}{SE(\hat{\beta})}$ is normally distributed with zero mean and unit variance. Therefore, we seek the smallest $n$ such that

$$f_n(\beta_1) := P \left( \left| \frac{\hat{\beta}}{SE(\hat{\beta})} \right| \geq 1.96 \middle| \beta = \beta_1 \right) \geq 0.8,$$

where $\beta_1$ is any number contained in the interval $[1, +\infty)$. Note here that $f_n(0) = 0.05$. Therefore,
\[ P \left( \left| \frac{\hat{\beta}}{SE(\beta)} \right| \geq 1.96 \mid \beta = \beta_1 \right) = P \left( \frac{\hat{\beta}}{SE(\beta)} \geq 1.96 \mid \beta = \beta_1 \right) + P \left( \frac{\hat{\beta}}{SE(\beta)} \leq -1.96 \mid \beta = \beta_1 \right) \]

\[ = P \left( \frac{\hat{\beta} - \beta_1}{SE(\beta)} \geq 1.96 - \frac{\beta_1}{SE(\beta)} \mid \beta = \beta_1 \right) \]

\[ + P \left( \frac{\hat{\beta} - \beta_1}{SE(\beta)} \leq -1.96 - \frac{\beta_1}{SE(\beta)} \mid \beta = \beta_1 \right) \]

\[ = P \left( Z \geq 1.96 - \frac{\beta_1}{SE(\beta)} \right) + P \left( Z \leq -1.96 - \frac{\beta_1}{SE(\beta)} \right). \]

The last line follows from the fact that if \( \beta = \beta_1 \), then \( \frac{\hat{\beta} - \beta_1}{SE(\beta)} \) follows a normal distribution with mean zero and unit variance. We denote by \( Z \) a random variable with this distribution.

By taking derivatives with respect to \( \beta_1 \) and using symmetry properties of the standard normal density, one can show that the expression

\[ P \left( Z \geq 1.96 - \frac{\beta_1}{SE(\beta)} \right) + P \left( Z \leq -1.96 - \frac{\beta_1}{SE(\beta)} \right) \]

is increasing for \( \beta_1 > 0 \), and thus

\[ f_n(1) = P \left( Z \geq 1.96 - \frac{1}{SE(\beta)} \right) + P \left( Z \leq -1.96 - \frac{1}{SE(\beta)} \right) \]

\[ \leq f_n(\beta_1), \]

for \( \beta_1 \) in \([1, +\infty)\).

Using the expression for \( SE(\hat{\beta}) \) obtained in part (b) and the prescribed value for \( \sigma \), we can write \( f_n(1) \) explicitly in terms of \( n \). It is enough to find the smallest \( n \) for which \( f_n(1) \geq 0.8 \). The best way to accomplish this is to write a program in R that iterates over values of \( n \) and stops when \( f_n(1) \geq 0.8 \). We find that \( n \) should be at least 93.